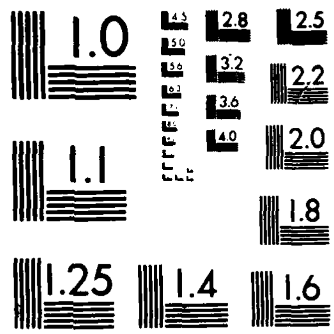
AD-A182 907   EVALUATING CREDIT APPLICATIONS A VALIDATION OF   1/1
MULTIATTRIBUTE UTILITY TE (U) UNIVERSITY OF SOUTHERN
AALIFORNIA LOS ANGELES SOCIAL SCIENCE R
UNCLASSIFIED   W G STILLWELL ET AL JUN 80 SSRI-RR-80-1   F/G 12/4   NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

AD-A182 907

0159 NOV 10 1980

social science research institute

UNIVERSITY OF SOUTHERN CALIFORNIA

RESEARCH REPORT

EVALUATING CREDIT APPLICATIONS: A VALIDATION OF MULTIATTRIBUTE UTILITY TECHNIQUES AGAINST A REAL WORLD CRITERION

William G. Stillwell, F. Hutton Barron, and Ward Edwards

Sponsored By:

Defense Advanced Research Projects Agency
Department of the Navy
Office of Naval Research

Monitored By:

Department of the Navy
Office of Naval Research
Pasadena Branch Office
Contract No. N00014-79-C-0038

June, 1980

SSRI RESEARCH REPORT 80-1

Social Science Research Institute
University of Southern California
Los Angeles, California 90007
(213) 743-6955

INSTITUTE GOALS:

The goals of the Social Science Research Institute are threefold:

- To provide an environment in which scientists may pursue their own interests in some blend of basic and methodological research in the investigation of major social problems.
- To provide an environment in which graduate students may receive training in research theory, design and methodology through active participation with senior researchers in ongoing research projects.
- To disseminate information to relevant public and social agencies in order to provide decision makers with the tools and ideas necessary to the formulation of public social policy.

HISTORY:

The Social Science Research Institute, University of Southern Californa, was established in 1972, with a staff of six. In fiscal year 1978-79, it had a staff of over 90 full- and part-time researchers and support personnel. SSRI draws upon most University academic Departments and Schools to make up its research staff, e.g. Industrial and Systems Engineering, the Law School, Psychology, Public Administration, Safety and Systems Management, and others. Senior researchers have joint appointments and most actively combine research with teaching.

FUNDING:

SSRI Reports directly to the Executive Vice President of USC. It is provided with modest annual basic support for administration, operations, and program development. The major sources of funding support are federal, state, and local funding agencies and private foundations and organizations. The list of sponsors has recently expanded to include governments outside the United States. Total funding has increased from approximately $150,000 in 1972 to almost $3,000,000 in the fiscal year 1978-1979.

RESEARCH INTERESTS:

Each senior SSRI scientist is encouraged to pursue his or her own research interests, subject to availability of funding. These interests are diverse; a recent count identified 27. Four major interests persist among groups of SSRI researchers: crime control and criminal justice, methods of dispute resolution and alternatives to the courts, use of administration records for demographic and other research purposes, and exploitation of applications of decision analysis to public decision making and program evaluation. But many SSRI projects do not fall into these categories. Most projects combine the skills of several scientists, often from different disciplines. As SSRI research personnel change, its interests will change also.

EVALUATING CREDIT APPLICATIONS: A
VALIDATION OF MULTIATTRIBUTE UTILITY TECHNIQUES
AGAINST A REAL WORLD CRITERION

Research Report 80-1

William G. Stillwell[1], F. Hutton Barron[2],
and Ward Edwards[1]

Social Science Research Institute,
University of Southern California[1]

and

The University of Kansas[2]

Sponsored By:

Defense Advanced Research Projects Agency

SUMMARY

Twenty-two credit officers from a major California lending institution served as subjects in a criterion validation of multiattribute utility elicitation techniques. The techniques tested were the Holistic Orthogonal Parameter Estimation (HOPE) technique (Barron and Person, 1978), Simple Multiattribute Rating Technique (SMART: Edwards, 1977), point distribution, and three rank weighting techniques as discussed in Stillwell and Edwards, 1979. Equal weighting of importance dimensions was also investigated. The criterion against which the judgments were compared was the lending institutions own credit scoring model. This model is based on statistical analysis of over 8,000 cases from the bank records and is a "best fit" prediction model.

Results demonstrate that subjective judgments of importance weighting show a high degree of agreement in application selection and in total utility realized from that selection. Decomposition techniques did somewhat better than holistic techniques.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This research was supported by the Advanced Research Projects Agency of the Department of Defense, and was monitored by the Office of Naval Research under Contract No. N00014-79-C-0038. The authors would like to thank the bank whose model and officers contributed to the quality of the study. They prefer to remain anonymous. In addition, we would like to thank Detlof von Winterfeldt and Richard S. John, whose comments on an earlier draft greatly contributed to the clarity and content of the final product.

DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

Introduction

Suppose that you are a bank officer and must decide whether credit should be granted to a number of applicants. With each credit application you receive data, for example, about the age and previous credit record of the applicant. You have some rules of thumb about how these data relate to overall creditworthiness, but there is too much information to integrate intuitively. How can you evaluate potential candidates in this situation?

Multi-attribute utility* measurement (MAUM) is the name of a class of models and measurement procedures developed to aid decision makers in such complex decision problems. MAUM evaluates options separately on each of a list of value relevant attributes. These single attribute evaluations are then combined by a formal model, usually using judgmental weights.

In the simplest case the weighted single attribute evaluations are added to obtain an overall value of the alternative. Formally, this additive model can be expressed as:

$$U(X(j) = \sum_{i=1}^{N} w(i)\ u(i)\ [x(ij)]$$

---

* The models tested in this paper are, strictly speaking, value models or riskless utility models. We simply refer to riskless utilities as utilities.

where U(x(j)) is the total utility for member (j) of option set x, w(i) is the weight for attribute i, and u(i) is the single dimension utility function transforming the value of x on dimension i into utility scaling. The additive model requires that attributes be preferentially independent (see Krantz, Luce, Suppes, and Tversky, 1971). Less formally, this means that the overall utility of an individual attribute is independent of other attribute values.

A number of methods have been proposed for determining both the u(i) functions and the w(i) weights. For practical purposes, these methods differ primarily in two ways: strength of theoretical justification and ease of use. Unfortunately, these two dimensions conflict. At one extreme are the highly complex, theoretically impeccable methods discussed by Keeney and Raiffa (1976) or Dyer and Sarin (1979). Somewhere in the middle are the easier but theoretically more problematic methods of Edwards's SMART technique (1977). Still simpler techniques can be based on ranking information (Stillwell and Edwards, 1979) or even equal weighting (Dawes and Corrigan, 1974). These simple techniques are defensible only as approximations--but that would be a highly persuasive defense if they led to essentially the same results as more complex and demanding methods.

This paper focuses on a comparison of weighting methods. Technical issues concerned with the u(i) can be equally important. But measures of u(i) have been less

controversial, since they are reasonably often simply monotone transformations on objective measures of i. In particular, the issue of central concern to this paper is whether or not complex and sophisticated methods of eliciting weights are worth while, in two different senses. Ultimately, a weighting method would be preferable to another in spite of additional difficulty in its use only if it did two things: changed the conclusion about what option is preferable, or by how much, and did so in a manner that made the conclusion more nearly correct.

Validity issues in MAU

The second of the two criteria mentioned above raises the most perplexing problem of any MAU technique: validity. Values are inherently subjective. In what sense, if any, can one elicitation technique be said to be more valid than another?

A familiar decision-analytic answer is: none. Most decision analysts apply the techniques as though validity, at least of utilities, is assumed. Practicing decision analysts, like other practitioners of clinical skills, must depend on user satisfaction as an important validating criterion. But if it is the only one, it is difficult to see how decision analysts are different from other well-trained and highly paid advisers who also give their clients satisfaction.

Aware of the difficulty, researchers have tried various approaches to validating decision-analytic tools and ideas.

A relatively traditional approach has depended on convergent validation (Pollack, 1964; Huber et. al, 1971; Fischer, 1971). This approach compares overall utilities calculated from a multi-attribute utility model (or statistically derived bootstrapping model) with holistic preference responses. MAUM utilities for each alternative are usually compared with holistic ratings over a set of alternatives or sometimes with choices among alternatives.

Other variations of the convergent approach compare results of different models and techniques with one another or even of different subjects with one another (see Fischer, 1977 for a more complete discussion of convergent validation in MAUM). Results of these and other studies of convergent validity have typically found correlations between decomposed and holistic responses of .7 to .9, with most in the high .80s to low .90s. Advocates of the convergent approach suggest that these results are "quite encouraging" (von Winterfeldt and Fischer, 1975). Shepard (1964), Hoepfl and Huber(1970), Edwards (1971) and others argue that MAUM procedures should not be validated using holistic responses as a criterion. Holistic responses may include substantial random error. (See Shepard, 1964; Slovic and Lichtenstein, 1971). Indeed, as Slovic, Fischhoff, and Lichtenstein (1977) point out, a decomposed judgment procedure that did capture the random as well as systematic components of holistic preferences would be indefensible as an improvement over the holistic procedure. Holistic responses may also

suffer from systematic bias. Responses may represent simplifying strategies of the decision maker. A more general argument also applies. If the goal of MAU procedures is to reproduce holistic judgments, they are a waste of time, since holistic judgments are usually easier to elicit.

The preceding paragraph is encouraging to defenders of MAU. Too high a correlation with holistic procedures would call the complexity of MAU procedures into question as unnecessary; too low a correlation would lead one to wonder whether the MAU procedures were in fact capturing the relevant values. Correlations in the .7 to .9 region are just about right for escaping both complaints.

Various procedures can be used to check whether the judgments that enter into a MAU elicitation conform to axioms of "reasonable behavior". Keeney and Raiffa (1976) spell out procedures for such tests, and Tversky (1967), von Winterfeldt (1971) and Fischer (1975) have studied conformity to various axioms experimentally. Such studies are usually not relevant to validity as here conceived. They test the appropriateness of specific axioms; if those axioms are inappropriate, the practicing decision analyst would face the viable options of ignoring the inappropriateness and treating the result as a good approximation (often an extremely useful and appropriate strategy) or of using other elicitation methods that do not depend on the violated axiom. Decision analytic elicitation

procedures exist in bewildering variety; failure of just about any axiom except the most fundamental ones (e.g. transitivity) can be circumvented.

While judgments that are consistent and orderly provide theoretical and practical justification of the MAUM model, no study of them can provide empirical demonstration of MAUM's ability to produce good decisions. A third approach to the validation problem therefore lies in finding an external criterion of correctness against which to validate value judgments. In the first such study, Yntema and Torgerson (1961) taught subjects the relationship between various cues and an arbitrary worth criterion. Then, using a number of different assessment procedures, they elicited the subjects' knowledge of the relationships. Since the experimenters had a priori defined the true relationships they could directly compare the subjects judgments with the results produced by the defined model.

The experimental procedure is essentially equivalent to the Brunswickian lens model paradigm (Brunswick, 1952; Hammond, 1966; Slovic and Lichtenstein, 1971) and its derivative, the Multiple Cue Probability Learning (MCPL) paradigm (Hammond, 1966; Slovic and Lichtenstein, 1971). In a MCPL study the subject is taught the relationship between individual cues and a criterion variable. For example, a subject could be taught a hypothetical relationship between the size, weight and speed of a football player and his overall ability. The relationship can and has been taught

by presenting feedback about the true model outcome (Schmitt, 1978), the true ratio of importance weights (Brehmer and Qvarnstrom, 1976), and/or validity coeffients themselves (Schmitt, Coyle, and Saari, 1977).

Although extensively used to examine subjective versus objective weighting techniques in the prediction context the MCPL paradigm has only recently come into use as a MAUM validation procedure (John and Edwards, 1978b; John, Collins and Edwards, 1980). In an experimental task in which subjects were asked to evaluate the dollar value of diamonds described on the four characteristics cut, color, carat, and clarity, subjects were taught an arbitrarily defined value model. As in the experiment discussed by Yntema and Torgerson (1961), various techniques were then used to elicit weight judgments. The model used to generate the training stimuli was thus a criterion against which to test the resulting judgments.

The results of this research argue for the use of MAUM techniques. In a recent review of importance weight assessment research, John and Edwards (1978a) conclude:

> "...the weighting literature reviewed, and particularly the recent criterion validation work, suggests that the concept of attribute importance is a psychologically meaningful one. For many of the laboratory and field settings studied, subjects gave

> responses to direct subjective assessments of importance weights that were both consistent (high convergent validity) and accurate (high criterion validity)."

Thus, at least in this highly contrived laboratory situation, subjects seem quite able to learn the relationships of individual cues to outcome variables and express these relationships in a meaningful, quantitative way. In addition, the work of John and Edwards (1978b) and John, Collins and Edwards (1980) provides direct evidence for subjects' ability to report what they have learned using standard MAUM techniques.

A clear picture emerges from the theoretical investigation of weighting judgment. Subjects in laboratory settings are able to both learn weighting functions and express them in response to MAUM elicitation techniques. But questions remain about whether decision makers in a real world setting perform equally well. In only a few cases has a real world criterion been used to evaluate the decomposition idea. Fischer (1977) discusses a study by Lathrop and Peters (1969) based on course evaluations for fourteen introductory Psychology courses. Students in those classes gave ratings of a number of individual factors for each course and an overall course evaluation rating. The ratings were averaged and the averages treated as objective value measures. Students who were not enrolled in these

courses either were given the average score of a class on each attribute and asked to judge the average overall rating (holistic judgment) or were simply asked to assign weights to each of the individual attributes (decomposed judgment). This study found that across a number of conditions, decomposed models afforded better prediction than did the intuitive judgments despite the fact that the subjective weights were decidedly non-optimal compared to weights derived from multiple regression.

A second study, performed by Eils and John (1980), again used college students as subjects. Groups of college student subjects were to evaluate potential credit applicants, described on 10 dimensions. The criterion was a statistically based credit model obtained from a local bank. The experimenters found that the SMART decomposition procedure (Edwards, 1977) significantly improved the ability of groups to produce judgments corresponding to the bank model criteria over holistic judgments.

The results of both studies support the decomposition approach. They are steps in the right direction. But the subjects were inexpert, the studies did not compare alternative weight elicitation techniques, and the only conclusion to which they can lead is that one decomposition procedure works better than an alternative based on holistic judgments.

This study sets out to remedy as many of these defects as possible. It uses highly expert subjects, performing a

task that they must perform every work day, and for which they are extensively trained. It uses a criterion that is both valid and realistic, in the sense that the procedures used to derive it ensure its validity, that decisions are based on it, and that the subjects are extensively trained on it and experienced in its use. The entire decision task is as realistic as an experiment permits; stimuli and issues bearing on the decisions are the same as in normal daily decision-making.

<u>Expertise and the Bank Model Criterion</u>

Most financial institutions use some statistical model to facilitate credit granting decisions for high volume, relatively low dollar amount loans, including decisions to give credit cards to would-be users. Many legal constraints limit the information the lending institution may use. Within these constraints, the credit scoring models use both readily available numbers and less structured inputs as predictors. For example descriptive attributes of credit applicants might include age, sex, credit history or even appearence.

One class of credit scoring model comes from applying discriminant analysis to good and bad accounts. Detailed definitions of "good" and "bad" vary from bank to bank; they depend on repayment history. Discriminant analysis finds the linear prediction equation that maximizes some difference measure between good and bad accounts, using weights on the available predictors.

Such a discriminant model was used as the criterion in this study. Its construction started with the collection of a sample of 4000 good and 4000 bad accounts, stratified by population and area. The analysis then determined which applicant attributes best discriminated between the good and bad accounts for this sample. It used the 7 best predictors in a percentage-of-variance-accounted-for sense. Table 1 shows the normalized weights for the bank model, ordered by rank. In addition, Table 1 shows the weight sets for rank sum, rank reciprocal and equal weights, weight approximation techniques to be discussed later.

The model thus derived was converted into an additive point scoring system for use by the bank officers as a decision aid. Each level of each attribute contributes points to a sum representing the creditworthiness of an applicant. Any point sum can be converted directly into a probability of default for that applicant.

Bank officers' experience with this point scoring system comes in several forms. First, the officers are given explicit model information. That is, they are told the exact relationship between the attribute levels and the overall credit score. In addition, they are explicitly trained in the relationship between attribute levels and the probability of default as determined from the sample data.

Bank officers also receive what is essentially outcome feedback from direct use of the model. As an application comes to the officer, that officer will first determine the

Table 1. Weight sets from normalized bank model and simplified techniques.

Dimension Rank	Bank Model	Rank Sum	Rank Reciprocal	Equal
1	.319	.250	.386	.143
2	.213	.214	.193	.143
3	.106	.179	.129	.143
4	.106	.143	.096	.143
5	.106	.107	.077	.143
6	.085	.071	.064	.143
7	.064	.036	.055	.143

overall credit score for the applicant based on information presented on the application. The officer then makes a credit decision for that applicant. The granting officer's name is then appended to the application and subsequent credit record. Tallies are kept over all applications approved by a given officer and an ongoing record is presented to that officer periodically. In addition, each time an account turns from good to bad the granting officer is given the entire credit file for review. Finally, the officers are given a monthly report in which the number of acceptances and rejections are broken down by credit score. There is, therefore, some pressure on the officer to avoid the simple strategy, ie. grant credit to only those applicants about whom there is relative certainty. It is interesting to note that bank records show that this bank extends credit to approximately 49% of its applicants.

The bank officers' experience with the model, in each of the forms discussed above, is extensive. During any given weeks' work an officer will make from 10 to over 1000 credit decisions to which the model is directly relevant. In addition, training in the use of the model and its relation to creditworthiness and probability of default is initially extensive and continues throughout the career of the officer.

The bank has a cut-off credit score at or above which extension of credit is recommended, below which the bank recommends that the application be rejected. This cut-off

score fluctuates periodically in response to the availability of money to the bank and the bank's financial condition. We must stress that this score is a recommendation only. The individual credit granting officer has a great deal of personal latitude for overriding the model recommendation. Of course, the amount and type of latitude is based on the record, position, and experience of the officer. Some officers may override the model with a simple signature, others must include an explanation while still others must convince another officer.

<u>Weight Elicitation Procedures</u>

<u>Rank Weighting Procedures</u>

Three different weight elicitation procedures were tested that use some aspect of the rank ordering of value dimensions to arrive at dimension weights. Two of the three require that the subject provide only the rank ordering of importance dimensions while the third requires the additional information of the weight assigned by the subject to the dimension considered most important. Each of these techniques is discussed in detail in Stillwell and Edwards (1979).

The first rank weighting procedure, called Rank Sum (RS) weighting is arrived at via the following formula:

$$W(i) = \frac{[1 - R(i) + 1]}{\sum_{j=1}^{N} [N - R(j) + 1]}$$

where W(i) is the normalized weight for dimension (i), N is the number of dimensions, and R(i) is the rank position of dimension (i). This rank weighting procedure is common in the weighting literature. Dimensions are simply given weight equivalent to the normalized inverse ranking of their place among other dimensions. For example, for a three dimension case the dimension ranked first would be given a weight of 3/(3+2+1)=.5.

Rank Reciprocal (RR) weights are derived from the normalized reciprocals of the dimension rank. They are defined by the following formula:

$$W(i) = \frac{1/R(i)}{\sum_{j=1}^{N} (1/R(j)}$$

where again W(i) is the normalized weight for dimension (i), R(i) is the rank of dimension (i) and N is the number of dimensions. For three dimensions the RR weight for the first dimension would be (1/1)/(1/1+1/2+1/3)=.55.

The third rank weighting procedure, Rank Exponent (RE) weights, requires one additional piece of information. The respondent judges the weight of the most important attribute on the usual 0-1 scale. Other weights are computed by:

$$W(i) = \frac{[N - R(i) + 1]^{z}}{\sum_{j=1}^{N} [N - R(j) + 1]^{z}}$$

where z is an exponent; the larger z is, the steeper the set of weights becomes. $z=1$ defines rank sum weights; $z=0$ defines equal weights. The other variables are the same as in equations (1) and (2). The respondent's judgment of $W(1)$ permits solution of the equation for Z, and given Z, the rest of the weights can be calculated. (See Stillwell and Edwards, 1979 for details).

Instructions for the rank ordering procedure asked respondents to put the attributes in order from most to least important in determining credit score. The point was stressed that attributes equal in importance should be indicated. The respondents were next asked to consider only the attribute they ranked first. They were to provide the proportion of the total weight that they would assign to that attribute.

Ratio Weighting

Three weight elicitation procedures result in weight sets said to have ratio properties. The first of these, the Simple Multi-Attribute Rating Technique (SMART) (Edwards, 1977) requires that the subjects first rank order the importance or value dimensions, then assign an arbitrary value of 10 to the dimension ranked last. Weights are then assigned to the other dimensions in ascending order, relative to the anchor weight on the lowest dimension, maintaining importance ratios between dimensions. For example, if the respondent considers the most important dimension 15 times as important as the least important one,

he or she should assign a weight of 150. The least important dimension is then disgarded, the second least important dimension given the value 10, and the ratio procedure repeated. At this point the respondent is asked to reconcile any inconsistencies. The SMART procedure followed the judgment of the weight to the most important dimension.

A second ratio weight elicitation procedure examined is called Holistic Orthogonal Parameter Estimation (HOPE), outlined in Barron and Person (1979). Essentially a Bootstrapping procedure (Slovic and Lichtenstein, 1971; Dawes, 1974), HOPE utilizes a fractionalized Analysis Of Variance (ANOVA) design to derive weights and location measures for categorical or categorized continuous variables. Subjects make a number of holistic judgments of decision alternatives determined by the design requirements. These judgments are analyzed via the ANOVA procedure whereby differences between marginal means are used as estimates of weights and location measures. For the purposes of this study, the HOPE procedure was constrained to an additive model. In order to conserve the respondents' time we were forced to provide an abbreviated HOPE design. All applications shown respondents included a single level of the attribute that had lowest weight in the bank model. This attribute could therefore not be evaluated since it had no variance. In addition, a single level was left out for two other attributes. Even with this shortened format

judgments of 25 applications were required of each respondent in a fractional design (Winer, 1971).

The holistic judgments required by the HOPE procedure were interspersed between each of the other sets of judgments. For each of the HOPE judgments respondents were presented with a single page on which appeared the attribute categories describing that application. A space was provided in which the subject was to give his or her judgment of the credit score for that applicant. We stressed to the officers that they were not to simply add up the scores for the individual attributes but instead give a judgment of the overall credit score.

In the final weight elicitation procedure subjects were asked to distribute 100 points over the value dimensions so as to reflect their feeling about the relative importance of value dimensions to total value (Hoffman, 1960). John and Edwards (1978a) suggest that this procedure leads subjects to attend to the differences between numbers of points given a pair of dimensions rather than the ratios. Although no empirical test of this suggestion has been made, if it is in fact true, the resulting weights could, at best, be treated as interval level information. The point distribution procedure followed the final set of HOPE judgments.

<u>Equal Weights</u>

In addition to the six weight elicitation techniques discussed above, equal weighting of importance dimensions was tested. Both experimental (Dawes and Corrigan, 1974)

and theoretical (Wainer, 1976; 1978) work have provided evidence and rationale for the effort saving device of simply adding the normalized single dimension utilities.

Method

Subjects

Subjects for the experiment were 22 officers from a major California bank. All respondents were familiar with the bank credit model used as criterion for their judgments and were experienced with making credit decisions as part of their normal job routine. Respondents ranged from 3 to 27 years (mean = 10.0) experience with credit lending institutions and from 1 to 27 years (mean=6.6) with their current employer.

Procedure

Each respondent was run in a single experimental session. These sessions ranged in length from 35 to 95 minutes. Each respondent worked individually with an experimentor. All experimenters had decision analytic training and experience.

Stimuli

Each respondent used a response booklet containing the total set of judgments required for all elicitation techniques. The order of presentation of weight elicitation procedures and location measure elicitation was partially determined by the nature of the information required. Location measure judgments were elicited before any of the weight elicitations were made so that respondents were aware of the ranges of the relevant attributes. Rank order weight elicitation judgments were made before ratio weight elicitations since SMART requires the rank order as input.

The order of presentation was thus:

- General instructions
- Respondent individual information
- Location measure judgments
- Ranking of attribute importance
- Weight of the most important attribute
- SMART judgments
- Point distribution
- The 25 holistic judgments required for HOPE were interspersed in a random order for each respondent, between other procedures.

Instructions stressed that we were trying to capture respondents' expertise in their judgments. They were told that they would make judgments both about individual applicants and about descriptive attributes of applicants for credit. We also asked subjects for general background information, (age, sex, etc.) and specific information about their credit granting experience (for example, years with this bank, number of credit models with which they have worked).

Respondents were next presented with a list of locations on or values of each attribute. They were asked to select the worst value of an attribute, assign a utility of 0, and then select the best value of the attribute and assign it a utility of 100. Respondents then placed the rest of the attribute values on this 0-100 scale relative to the endpoints. This procedure constituted the location

measure elicitation. Finally, respondents made weight elicitation judgments as discussed earlier.

Respondents completed the judgments necessary for each procedure before going on to the next. Respondents were asked not to refer back to previous judgments or change any of those judgments. All elicitations were done interactively until the experimenter was confident that the subject understood the procedure. Questions were allowed at any time during the experimental session and subjects were encouraged to express any confusion or misunderstanding. Our hope was to examine the procedures in a form as near as possible to that in which they would be found in a real world application of that technique.

Results

The data analysis for this experiment is in two parts. First, we directly compared the normalized weight sets that resulted from respondents' judgments. Two such comparisons were made. Table 2 shows the true weight and the mean, median, and standard deviation, across respondents, of weights from each attribute by each elicitation technique. Attributes are numbered in order of true weight. Looking across attributes several things become evident. In each of the self-explicated weighting techniques, both median and mean responses show that respondents felt attribute 2 to be more important than attribute 1. But the weights derived from the holistic judgments of HOPE suggest that when actually making judgments of credit score respondents correctly identify attribute 1 as more important. A second finding is that SMART and rank exponent weighting result in more peaked weight sets as evidenced by the larger ratios between the highest and lowest weighted attributes. HOPE cannot be so evaluated. For HOPE, the lowest weighted attribute was not included in the design and thus, this ratio has no meaning. Finally, the results suggest that although analysis of holistic responses correctly identified the most important attribute, the rest of the attributes are very close in mean and median weight. On the other hand, the self-explicated techniques correctly produced weights for the first two attributes that are much higher than for the attributes ranked third thru seventh.

Table 1. Mean, Standard Deviation and Median Weight Judgments Across Subjects

Attribute	True	Hope	Smart	RS	RR	RE	100 PTs
1							
$\bar{X}$	.319	.253	.256	.212	.227	.272	.226
S.D.	----	.167	.145	.035	.106	.151	.149
M.D.	----	.231	.227	.214	.193	.256	.235
2							
$\bar{X}$	.214	.169	.299	.228	.293	.326	.264
S.D.	----	.062	.178	.033	.108	.174	.077
M.D.	----	.235	.246	.250	.386	.275	.250
3							
$\bar{X}$	.106	.124	.110	.143	.103	.102	.123
S.D.	----	.039	.053	.033	.028	.053	.040
M.D.	----	.110	.124	.143	.096	.116	.130
4							
$\bar{X}$	.106	.157	.070	.089	.074	.049	.084
S.D.	----	.025	.034	.038	.018	.051	.037
M.D.	----	.151	.061	.071	.065	.028	.100
5							
$\bar{X}$	.106	.156	.142	.165	.134	.139	.154
S.D.	----	.052	.063	.041	.069	.072	.045
M.D.	----	.142	.145	.179	.129	.139	.150
6							
$\bar{X}$	.085	.112	.089	.111	.108	.088	.104
S.D.	----	.049	.078	.062	.095	.104	.069
M.D.	----	.099	.081	.071	.077	.046	.100
7							
$\bar{X}$	.064	.000	.037	.052	.061	.024	.048
S.D.	----	.000	.043	.034	.016	.040	.043
M.D.	----	.000	.022	.036	.055	.014	.047
Ratio of highest weighted attribute to lowest	4.98	----	8.08	6.33	4.30	13.6	5.5

In order to analyze more closely the quality of weight judgments a second comparison of the weights resulting from the different elicitation techniques was performed. We examined Cumulative Frequency Distributions (CFD) of the absolute values of the differences between the true weights and those resulting from each elicitation procedure and approximation technique. This analysis was across both subjects and dimensions. For this analysis we define dominance in a CFD as: CFD A dominates CFD B if and only if for any value of absolute difference the cumulative frequency for A is greater than or equal to the cumulative frequency for B.

Only a few distributions show dominance over the entire range of values. The difference distribution of rank sum weights dominates those of rank exponent, SMART, and equal weights. HOPE dominates rank exponent and equal weights and point distribution dominates equal weights. In terms of the average absolute deviation the ordering of techniques is rank sum (49.5), point distribution (51.2), HOPE (52.4), rank reciprocal (56.3), SMART (69.9), Equal weights (70.7), and rank exponent (79.6).

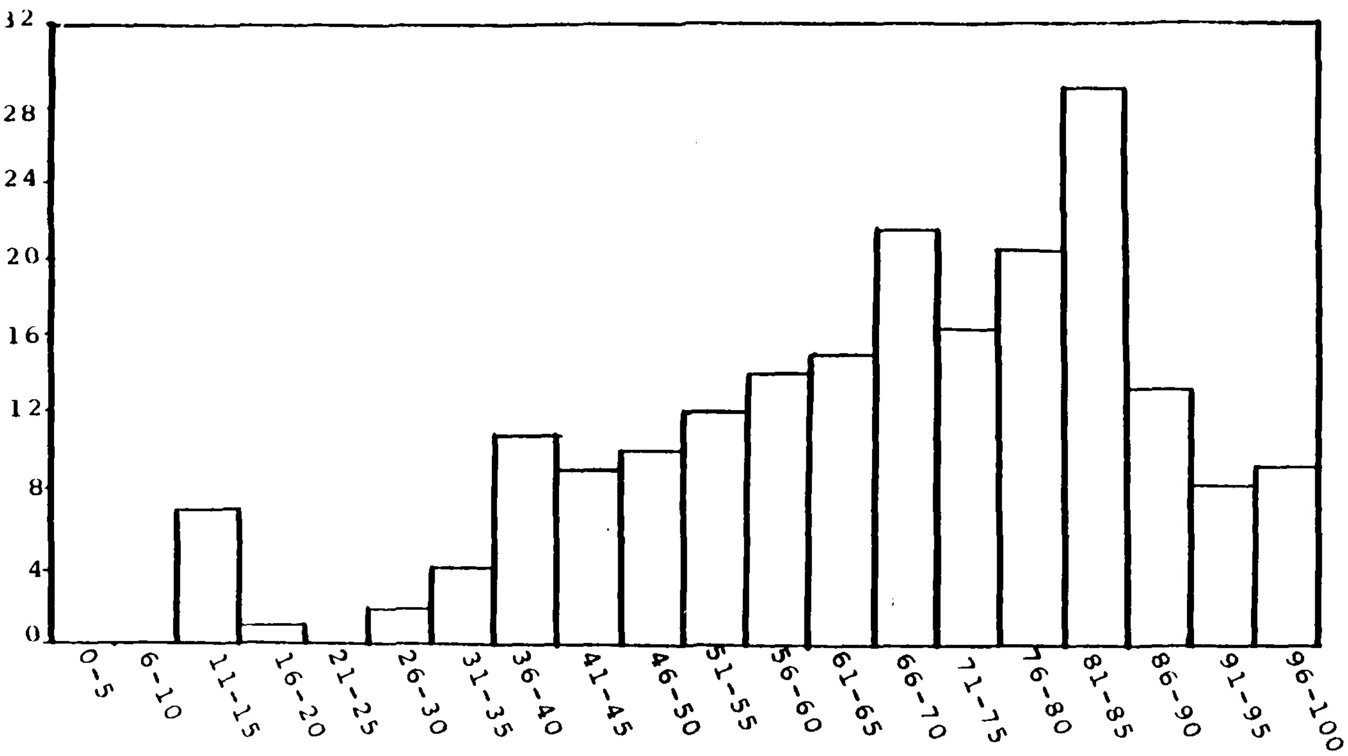
The second part of our analysis addresses the practical significance of the differences found in weight judgments. We looked at the same type of decisions the bank officers make in the performance of their job. For this purpose we used a sample of 200 real applications for credit at the bank. These applications were chosen to be representative

of the general population of applications that an officer is likely to see in his or her usual job performance. Figure 1 displays the distribution of true utilities of these 200 applications as calculated from the bank model. It is apparent that the distribution of true utilities (rescaled from 0 to 100) is skewed slightly to the left. The mean of the distribution is 66.3. A value of 68.1 is the decision point equal to or above which credit is given as outlined by bank rules.

Substantial negative correlations between attributes in a multi-attributed context can lead to weight sensitivity and, in the presence of suboptimal weights, poor selection ordering (Stillwell and Edwards, 1979; Newman, 1977; Newman, Seaver and Edwards, 1976; McClelland, 1978). Table 3 shows the correlations between dimensions for the 200 sample applications. No correlation is meaningfully negative. This fact guarantees that all weighting procedures, including equal weights, will do reasonably well. One handicap of the quest for realism in stimuli, criteria, and respondents is that we must take the stimuli we can get, and cannot design into them properties that would increase the strength of the experimental design. Even if we could have designed negative correlation into the applicant set, we would have hesitated to do so. The resulting applicant set would inevitably have seemed very strange indeed to the respondents.

In order to compare elicitation procedures, values of

Figure 1. Distribution of bank model utilities for 200 sample applications.



$\bar{X} = 66.3$

Table 3. Interdimensional correlations: 200 sample applications.

	D1	D2	D3	D4	D5	D6
D2	.047					
D3	.219	.158				
D4	.209	.008	.407			
D5	.197	.389	.367	.373		
D6	.044	.255	.323	.333	.344	
D7	.133	.126	.022	-.006	.148	.055

overall utility were calculated for each of the 200 applications using the bank model and each of the weight sets from the different elicitation procedures and location measure sets. For each subject the utilities derived from the bank model were then correlated with those calculated from each of the weight elicitation procedure-location measure combinations. These correlations were then averaged across subjects. The results of this analysis are shown in Table 4. For example, the average correlation, across 22 subjects, between overall utilities calculated from the bank model and those from the SMART weight elicitation procedure and judgmental location measures is .881.

The bank credit scoring model led to the selection of 98 of the 200 applications for credit. In addition to the correlations, Table 4 shows the average number out of those 98 that would have been chosen by each of the other techniques. For instance, using HOPE weights and HOPE location measures, an average across subjects of 77.8 of the correct 98 would have been granted credit. Assuming that 98 applicants were to be extended credit this also means, of course, that an average of 20.2 applications would have been given credit by the HOPE procedure that would not have been given credit by the bank model.

The last column of Table 4 shows the proportion of total utility, as calculated by the bank model, that would have been realized from selections resulting from each set-location measure combination. Again this assumes that 98

Table 4. Correlation, correct number selected, and total utility comparisons between each weight procedure-location measure combination and the bank model.

Weight tech.	Location meas. set	Avg. corr.	Average # of apps. in top 98 as selected by bank model	Proport[illegible] utility Capture[illegible] Max(98)[illegible] Min(98)[illegible]
HOPE	HOPE	.730*	77.8*	.883*
SMART	Judgmental	.881	85.3	.953
Rank Sum	Judgmental	.934	87.7	.965
Rank Recip.	Judgmental	.887	83.9	.9[illegible]
Rank Exponent	Judgmental	.860	82.6	.9[illegible]
Dist. 100 Pts.	Judgmental	.921	86.5	.9[illegible]
Equal	Judgmental	.926	86.0	.95[illegible]
SMART	Bank model	.923	88.8	.9[illegible]
Rank Sum	Bank model	.964	91.2	.9[illegible]
Rank Recip.	Bank model	.927	88.5	.9[illegible]
Rank Exponent	Bank model	.907	86.8	.97[illegible]
Dist. 100 Pts.	Bank model	.959	90.6	.9[illegible]
Equal	Bank model	.938	86.0	.90[illegible]

*One subject is not included in this average. Due to inappropriate responses to the holistic judgments no HOPE weights or location measures could be calculated.

were to be granted credit. This number is scaled such that 1.0 is the total utility of the best 98 applications as determined by the bank model and 0.0 is the total utility of the lowest 98. For example, if the decision maker had used rank reciprocal weights and the bank model location measures, the 98 selections, averaged across subjects, would have resulted in 97% of the total possible utility being realized.

The findings expressed in Table 4 are relatively consistent across the three analyses so we will discuss them together. First, and by far most important, is the fact that all procedures do remarkably well. Except for the HOPE procedure, all average correlations are above .86, more than 82.6 out of 98 applications were selected correctly for each weight set-location measure combination, and a minimum of 93.5% of the total possible utility was realized. Given that all techniques perform near the maximum, it is virtually impossible to differentiate between them on the basis of aggregate performance indices. Still, some qualified statements can be made. There is some indication of sensitivity to error in location measure judgments. We found that approximately 30% of dimensions had non-monotonicities for the judgmental location measures leading to an average drop in correlation from the bank model location measures of .035. A drop of 3.25 was found in the number of applications correctly identified as worthy of credit leading to a drop of 2% in the total utility

captured. The HOPE procedure resulted in very good weight judgments but suffered most from poor location measure placement. Over 78% of HOPE dimensions had non-monotonic category placement.

A second interesting finding is the quality of the performance of equal weighting of importance dimensions. In agreement with the theoretical findings of Wainer (1976; 1978) and Einhorn and Hogarth (1975) we found that simple equal weighting of importance dimensions provided a remarkably good approximation to the weighting of the true bank model.

Discussion

Expert subjects used several well known multi-attribute utility weight elicitation techniques. The purpose of this experiment was to find out how well each of these assessment techniques replicated the results of a criterion model developed in the environment of subjects' expertise. Both the normalized weights and the decisions produced by the weights were used for the comparison.

The use of judgmental decomposition methods to assess multi-attribute utilities for credit applicants in this study led to the same high quality of decisions found in previous studies (Lathrop and Peters, 1969; John and Edwards, 1979; John, Collins and Edwards, 1980). Although there seem to be differences in the quality of the weights themselves from one technique to another, these differences do not pass along to the resulting decisions. There was very little difference between the elicitation procedures in the quality of these decisions and, in fact, simple equal weighting of attributes performed extremely well.

The results of a holistic, bootstrapping procedure were generally poorer. These results conflict with previous studies of this technique (Barron and Person, 1979; John, Collins and Edwards, 1980) as well as more general work on holistic judgment (see, for example, Fischer, 1977; Dawes and Corrigan, 1974). The reasons for this poorer performance are not altogether clear, but it seems likely that changes from the experts' normal judgment situation

dictated by time and technique constraints led to at least a part of this decrement.

When making credit decisions in the performance of their duties, the experts generally make a simple dichotomous decision, ie. credit-no credit. Only those decisions very near the cut-off score require serious consideration in this type of judgment. Those much higher or lower need only cursory examination before the decision becomes obvious. The HOPE procedure relies on judgments across the range of value on all attributes such that many of the holistic judgments required were some distance from the cut-off score. The experts are not experienced at close consideration of these judgments and poor judgments of these extreme values could account for our results.

All procedures other than HOPE produced decisions of such high quality that, so far as these data can guide us, the appropriate basis for weighting judgments is ease of use. We do not argue for the generality of this conclusion-especially as it might be applied to negatively correlated values.

The major difference found between the self-explicated weighting procedures and the holistic procedure needs further investigation. The difference may be due to the task environment. Knowledge of the model is made available to the experts, knowledge very similar to that required by the decomposition procedures, while their "holistic expertise" was limited to categorical judgments (accept,

reject). Unfortunately, 20 of the 25 cases used to elicit the holistic judgments were easily classified as "reject". This may have severely affected the accuracy of the required holistic rating judgment.

Another reason for this finding may lie in the attributes themselves. Attribute 1, the most important predictor, includes historical information, while attribute 2 is purely a measure of immediate situation. In decomposed judgments, the respondents may have given most weight to the obviously important attribute that best describes the current state of the applicant, while in holistic judgments they may have assumed that relevant history incorporates situational information. (We regret that the requirement to keep the attributes confidential precludes a more detailed discussion of the point.)

It is important to note the similarity of our results with those of the MCPL study discussed earlier. John, Collins and Edwards (1980) found high convergence between a number of subject weight elicitation techniques and the criterion, as was found in this study. The implication for future work is obvious. We can, with confidence, extend the MCPL studies to investigation of real world situations where no criterion exists.

Finally, a note of caution must be introduced. As discussed earlier, the nature of the applications seen by the bank officers, where all attributes were positively related, makes this an insensitive situation for the

comparison of multi-attribute utility elicitation techniques. We cannot be certain whether in another, more sensitive decision situation strong differences would have been found. In addition, we cannot estimate the ubiquity of this insensitive situation for decision makers. Our results merely show that in a single real world decision situation experts are able to produce quality decisions using a number of decomposition procedures. Our findings do not make meaningful sensitivity analyses for important decision problems unnecessary or irrelevant.

Reference Notes

1. Fischer, G. W. Four methods for assessing multi-attribute utilities: An Experimental validation. Technical Report Nr. 037230-6-T, Engineering Psychology Laboratory, University of Michigan, 1972.

2. John, R. S. & Edwards, W. Importance weight assessment for additive, riskless preference functions: A Review. Research Report 78-5, Social Science Research Institute, University of Southern California. Los Angeles, 1978a.

3. John, R. S. & Edwards, W. Subjective statistical importance weights: A criterion validation. Research Report 78-7, Social Science Research Institute, University of Southern California, Los Angeles, 1978b.

4. John, R. S., Collins, L. & Edwards, W. A Comparison of importance weights for MAUA derived from holistic, indifference, direct subjective and rank order judgments. Technical Report 80-4, Social Science Research Institute, University of Southern California, 1980.

5. Keeney, R. L. Energy policy and value tradeoffs. Research Memorandum 75-76, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1975.

6. Lathrop, R. G. & Peters, B. E. Subjective cue weighting and decisions in a familiar task. Preceedings, 77th Annual Convention, APA, 1969.

7. Newman, J. R., Seaver, D. A. & Edwards, W. Unit versus differential weighting schemes for decision making: A method of study and some preliminary results. Technical Report 76-5, Social Science Research Institute, University of Southern California, 1976.

8. Raiffa, H. Preferences for multi-attributed alternatives (RM-5868-DOT/RC). The Rand Corporation, April, 1969.

9. Stillwell, W. G. & Edwards, W. Rank weighting in multi-attribute utility decision making: Avoiding the pitfalls of equal weight. Research Report 79-2, Social Science Research Institute, University of Southern California, Los Angeles, 1979.

10. von Winterfeldt, D. Multi-attribute utility theory: Theoretical background and an experimental validation. The University of Michigan, 1971.

References

Barron, F. H. & Person, H. B. Assessment of multiplicative utility functions via holistic judgments. OBHP, 1979, 24, 147-166.

Brehmer, B & Qvarnstrom, G. Information integration and subjective weights in multiple-cue judgments. OBHP, 1976, 17, 118-126.

Brunswik, E. The Conceptual framework of Psychology. Chicago: University of Chicago Press, 1952.

Dawes, R. M. A case study of graduate admission: Application of three principles of human decision making. American Psychologist, 1971, 26, 180-187.

Dawes, R. M. & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Dyer, J.S. and Sarin, R.A. Measurable multiattribute value functions. Operations Research, 1979, 27, 810-822.

Eckenrode, R. T. Weighting multiple criteria. Management Science, 1965, 12, 180-191.

Edwards, W. Social utilities. In Decision and Risk Analysis: Powerful New Tools for Management, Proceedings of Sixth Triannual Symposium, June 1971, Hoboken: The Engineering Economist, 1972, 119-129.

Edwards, W. How to use multi-attribute utility measurement for social decision making. In IEEE Transactions on Systems, Man, and Cybernetics. March 1977. 326-340.

Eils, L. C. III & John, R. S. A Criterion validation of multi-attribute utility analysis and of group communication strategy. OBHP, 1980, 25, 268-288.

Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. OBHP, 1971, 6, 1-27.

Einhorn, H. J. & Hogarth, R. Unit weighting schemes for decision making. OBHP, 1975, 13, 171-192.

Ewen, R. B. Weighting components of job satisfaction. J. of Applied Psychology, 1967, 51, 68-73.

Fischer, G. W. Experimental application of multiattribute utility models. In Wendt and Vlek (Eds.), Utility, Probability, and Human Decision Making. Dordrecht, Holland: Reidel. 1975.

Fischer, G. W. Convergent validation of decomposed multi-attribute utility assessment procedures for risky and riskless decisions. OBHP, 1977, 18, 295-315.

Hammond, K. R. Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of Psychology. In K. R. Hammond (Ed.), The Psychology of Egon Brunswik. New York: Holt, Rinehart, & Winston, 1966.

Hoepfl, R. T. & Huber, G. A. A study of self-explicated utility models. Behavioral Science, 1970, 15, 408-414.

Hoffman, P. J. The Paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 57, 116-131.

Huber, G. P., Daneshgar, R. & Ford, D. L. An Empirical comparison of five utility models for predicting job preferences. OBHP, 1971, 6, 267-282.

Huber, G. P., Sahney, V. K. & Ford, D. L. A Study of subjective evaluation models. Behavioral Science, 1969, 14, 483-489.

Keeney, R. L. & Raiffa, H. Decisions with Multiple Objectives. Preferences and Value Tradeoffs. John Wiley and Sons, 1976.

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. Foundations of Measurement, Volume 1. Academic Press, New York. 1971.

McClelland, G. Equal versus differential weighting for multi-attribute decision: There are no free lunches. Psychological Bulletin, 1980, 87, (in press).

Mikes, P. S. & Hulin, C. L. Use of importance as a weighting component of job satisfaction. J. of Applied Psychology, 1968, 52, 394-398.

Newman, J. R. Differential weighting in mulit-attribute utility measurement: When it should and when it should not make a difference. OBHP, 1977, 20, 312-324.

Pollack, I. Action selection and Ynetma-Torgerson worth function. In Information system science and engineering: Proceedings of the first congress of the information systems sciences, New York: McGraw-Hill, 1964.

Pollak, R. A. Additive von Neumann-Morgenstern utility functions, Econometrica, 1967, 35, 485-495.

Schmitt, N. Comparison of subjective and objective weighting strategies in changing task situations. OBHP, 1978, 21, 171-188.

Schmitt, N., Coyle, B. W., & Saari, B. B. Types of task information feedback in multiple-cue probability learning. OBHP, 1977, 18, 316-328.

Shepard, R. N. On subjectively optimum selection among multi-attribute alternatives. In M.W. Shelley and G. L. Bryan (Eds.), Human Judgment and Optimality. New York: Wiley, 1964.

Slovic, P. Analyzing the expert judge: A descriptive study of a stockbroker's decision process. Journal of Applied Psychology, 1969, 53, 255-263.

Slovic, P., Fischhoff, B., & Lichtenstein, S. C. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Slovic, P. & Lichtenstein, S. C. Comparison of Bayesian and regression approaches to the study of information processing in judgment. OBHP, 1971, 6, 649-744.

Tversky, A. Additivity, utility, and subjective probability. J. of Mathematical Psychology, 1967, 4, 175-202.

von Winterfeldt, D. & Fischer, G. W. Multi-attribute utility theory: Models and assessment procedures. In Wendt and Vlek (Eds.) Utility, Probability, and Human Decision Making. Dordrecht, Holland: Reidel, 1975.

Wainer, H. Extimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Wainer, H. On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.

Winer, B. J. Statistical Principles in Experimental Design, 2nd Edition. McGraw-Hill: New York, 1971.

Yates, J. F. & Jagacinski, C. M. Reference effects in multi-attribute evaluations. OBHP, 1979, 24, 400-410.

Yntema, D. B. & Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, HFE- 2, 20-26.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 001595	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Evaluating credit applications: A validation of multiattribute utility techniques against a real world criterion		5. TYPE OF REPORT & PERIOD COVERED Technical -
		6. PERFORMING ORG. REPORT NUMBER 80-1
7. AUTHOR(s) W.G. Stillwell, F.H. Barron, and W. Edwards		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0038
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research - Dept of the Navy - 800 Quincy Street Arlington, VA 22217		12. REPORT DATE June, 1980
		13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Research Branch Office 1030 East Green Street Pasadena, California 91101		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiattribute utility
credit scoring, validation, elicitation, subjective weighting, decomposition, creditworthiness, criterion

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Twenty-two credit officers from a major California lending institute served as subjects in a criterion validation of multiattribute utility elicitation techniques. The techniques tested were the Holistic Orthogonal Parameter Estimation (HOPE) technique (Barron and Person, 1977), Simple Multiattribute Rating Technique (SMART: Edwards, 1977), point distribution, and three rank weightin techniques as discussed in Stillwell and Edwards, 1979. Equal

weighting of importance dimensions was also investigated. The criterion against which the judgments were compared was the lending institutions own credit scoring model. This model is based on statistical analysis of over 8,000 cases from the bank records and is a "best fit" prediction model.

Results demonstrate that subjective judgments of importance weighting show a high degree of agreement in application selection and in total utility realized from that selection. Decomposition techniques did somewhat better than holistic techniques.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense Research and Engineering (Test and Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
Camp H. M. Smith, Hawaii 96861

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. [illegible]
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Cent[illegible]
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Oper[illegible]
(OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology [illegible]
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research [illegible]
800 North Quincy Street
Arlington, VA 2221[illegible]

Office of Naval Research
Naval Analysis Programs (Code [illegible]
800 North Quincy Street
Arlington, VA 2221[illegible]

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Attention: R. J. Roland, (Code 52
C^3 Curriculum
Monterey, CA 93940

Naval Personnel Research and Developmen
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N2
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Rese...
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
Research Laboratory
U.S. Army Institute for the Behavioral
and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts
Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
(Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
and Acquisition Programs
Office of the Deputy Chief of Staff for
Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and
Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, DC 20330

Director, Advanced Systems Division
(AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phill...
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Comp... Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 229[illegible]

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norman, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309

Decisions and Designs, Incorporated
Suite 600, 8400 Westpark Drive
P.O. Box 907
McLean, Virginia 22101

END

9-87

DTIC